Edinburgh Research Explorer

Ovarian volume throughout life: a validated normative model

Citation for published version:

Digital Object Identifier (DOI):
10.1371/journal.pone.0071465

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
PLoS ONE

Publisher Rights Statement:
Copyright: © 2013 Kelsey et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Ovarian Volume throughout Life: A Validated Normative Model

Thomas W. Kelsey¹, Sarah K. Dodwell², A. Graham Wilkinson³, Tine Greve⁴, Claus Y. Andersen⁴, Richard A. Anderson⁵, W. Hamish B Wallace⁶*

¹ School of Computer Science, University of St Andrews, St Andrews, Fife, United Kingdom, ² School of Medicine, University of Edinburgh, Edinburgh, United Kingdom, ³ Department of Paediatric Radiology, Royal Hospital for Sick Children, Edinburgh, United Kingdom, ⁴ Laboratory of Reproductive Biology, Section 5712, The Juliane Marie Centre for Women, Children and Reproduction, University Hospital of Copenhagen, University of Copenhagen, Copenhagen, Denmark, ⁵ MRC Centre for Reproductive Health, Queens Medical Research Institute, University of Edinburgh, Edinburgh, United Kingdom, ⁶ Department of Haematology/Oncology, Royal Hospital for Sick Children, Edinburgh, United Kingdom

Abstract

The measurement of ovarian volume has been shown to be a useful indirect indicator of the ovarian reserve in women of reproductive age, in the diagnosis and management of a number of disorders of puberty and adult reproductive function, and is under investigation as a screening tool for ovarian cancer. To date there is no normative model of ovarian volume throughout life. By searching the published literature for ovarian volume in healthy females, and using our own data from multiple sources (combined n = 59,994) we have generated and robustly validated the first model of ovarian volume from conception to 82 years of age. This model shows that 69% of the variation in ovarian volume is due to age alone. We have shown that in the average case ovarian volume rises from 0.7 mL (95% CI 0.4–1.1 mL) at 2 years of age to a peak of 7.7 mL (95% CI 6.5–9.2 mL) at 20 years of age with a subsequent decline to about 2.8 mL (95% CI 2.7–2.9 mL) at the menopause and smaller volumes thereafter. Our model allows us to generate normal values and ranges for ovarian volume throughout life. This is the first validated normative model of ovarian volume from conception to old age; it will be of use in the diagnosis and management of a number of diverse gynaecological and reproductive conditions in females from birth to menopause and beyond.

Introduction

The main functions of the ovary are to provide gametes and sex steroids to allow and support the establishment of pregnancy, and act as a repository for the non-growing follicles (NGFs) that allow this process to take place over several decades. The main constituents of the ovary are therefore its follicle endowment (both growing and non-growing), and the stromal tissues that support these functions. The human ovary establishes its complete complement of non-growing follicles during fetal life, and after birth there is a continuous process of recruitment until menopause at an average age of 50–51 years, when fewer than one thousand remain [1–3]. There is a wide variation in the age at menopause between individuals [4,5] and it is thought that this is due in large part to variations in the initial endowment of NGFs [1]. Currently, clinical assessment is unable to assess reliably the number of NGFs, or their rate of loss or activation.

Ovarian volume is one of several putative biomarkers of the ovarian reserve, others include serum anti-Müllerian Hormone (AMH), and antral follicle count (AFC) which have been shown to have clinical utility in the assessment of women with subfertility [6]. Ovarian volume is currently one of the diagnostic criteria for the most common endocrinopathy in women (polycystic ovary syndrome; PCOS) [7,8] and may be of value in screening for ovarian cancer [9]. We have shown a strong and positive correlation between ovarian volume and NGF population in the human ovary for ages 25–51 years [10], but there is only sparse information available on ovarian volume in healthy young girls and women [11]. A greater understanding of the changes in ovarian volume throughout life are likely to be helpful in the diagnosis and treatment of many disorders in gynaecology and reproductive medicine [12].

The data on ovarian volume in young girls is limited due to the lack of an easy non-invasive method of imaging the ovaries accurately. Much of the data that is published is in girls with abnormalities in pubertal development and so does not reflect the healthy population [13,14]. In the adult woman the advent of transvaginal ultrasound as a routine gynaecological technique has led to a large source of data on ovarian volume in healthy women [15]. To date no single study has examined ovarian volume across the lifespan in healthy females. The aim of this study is to develop a validated model of ovarian volume in healthy females from conception throughout life from data aggregation from multiple sources.
Results

The validated model is a degree 14 polynomial of the form

\[ \log_{10}(\text{Volume} + 1) = c_0 + c_1 \times \text{age} + c_2 \times \text{age}^2 + \ldots + c_{14} \times \text{age}^{14} \]

with coefficients \( c_i \) given in Table 1, and relationship to the data given in Figure 1. The model has coefficient of determination \( r^2 = 0.69 \) indicating that around 69% of the variation in human ovarian volumes is due to age alone. The residual plot (Figure 2) shows a distribution close to the ideal Gaussian curve \( (r^2 = 0.993) \), this coefficient of determination being higher than that for three other possible curves for these residuals. Moreover, the proportions of residuals within one, two and three standard deviations (respectively 71%, 96% and 99%) are close to the expected values for data with a Gaussian distribution (respectively 68%, 95% and 99%). Figure 3 is an exemplar of the 5-fold validation process in which a model is chosen that neither overfits nor underfits the underlying dataset.

The log-unadjusted predictive normative model is shown in Figure 4. This shows the mean volume per ovary in millilitres (mL) for the healthy human population, together with prediction intervals at ±1 and ±2 standard deviations (SD). Approximately 68% of ovarian volumes are expected to lie within ±1 SD of the mean; approximately 95% within ±2 SD of the mean. Mean and normative ranges for ovarian volumes are given for ages from birth to 50 years in Table 2. Our model shows that in the average case ovarian volume rises from 0.7 mL (95% CI 0.4–1.1 mL) at 2 years of age to a peak of 7.7 mL (95% CI 6.5–9.2 mL) at 20 years of age

![Figure 1. The validated model of log-adjusted ovarian volume throughout life.](image)

The \( r^2 \) coefficient of determination indicates that 69% of the variation in human ovarian volumes is due to age alone. Colour bands indicate ranges within ±1 standard deviation from mean, within ±1 and ±2 standard deviations, and outside 2 standard deviations.

doi:10.1371/journal.pone.0071465.g001

Table 1. Coefficients for the validated model.

<table>
<thead>
<tr>
<th>( i )</th>
<th>Coefficient ( c_i )</th>
<th>Standard Error</th>
<th>T Value</th>
<th>95% Conf Lim</th>
<th>95% Conf Lim</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8.92E-02</td>
<td>8.00E-03</td>
<td>11.2</td>
<td>1.24E-02</td>
<td>1.91E-01</td>
</tr>
<tr>
<td>1</td>
<td>1.10E-01</td>
<td>8.28E-03</td>
<td>13.3</td>
<td>5.22E-03</td>
<td>2.16E-01</td>
</tr>
<tr>
<td>2</td>
<td>-3.05E-02</td>
<td>5.92E-03</td>
<td>-5.2</td>
<td>-1.06E-01</td>
<td>4.47E-02</td>
</tr>
<tr>
<td>3</td>
<td>5.09E-03</td>
<td>1.63E-03</td>
<td>3.1</td>
<td>-1.56E-02</td>
<td>2.58E-02</td>
</tr>
<tr>
<td>4</td>
<td>-4.35E-04</td>
<td>2.34E-04</td>
<td>-1.9</td>
<td>-3.40E-03</td>
<td>2.53E-03</td>
</tr>
<tr>
<td>5</td>
<td>2.49E-05</td>
<td>2.03E-05</td>
<td>1.2</td>
<td>-2.33E-04</td>
<td>2.83E-04</td>
</tr>
<tr>
<td>6</td>
<td>-1.23E-06</td>
<td>1.15E-06</td>
<td>-1.1</td>
<td>-1.59E-05</td>
<td>1.34E-05</td>
</tr>
<tr>
<td>7</td>
<td>5.43E-08</td>
<td>4.49E-08</td>
<td>1.2</td>
<td>-5.17E-07</td>
<td>6.25E-07</td>
</tr>
<tr>
<td>8</td>
<td>-1.89E-09</td>
<td>1.23E-09</td>
<td>-1.5</td>
<td>-1.75E-08</td>
<td>1.38E-08</td>
</tr>
<tr>
<td>9</td>
<td>4.66E-11</td>
<td>2.40E-11</td>
<td>1.9</td>
<td>-2.59E-10</td>
<td>3.52E-10</td>
</tr>
<tr>
<td>10</td>
<td>-7.87E-13</td>
<td>3.31E-13</td>
<td>-2.4</td>
<td>-4.99E-12</td>
<td>3.42E-12</td>
</tr>
<tr>
<td>11</td>
<td>8.87E-15</td>
<td>3.15E-15</td>
<td>2.8</td>
<td>-3.12E-14</td>
<td>4.89E-14</td>
</tr>
<tr>
<td>12</td>
<td>-6.36E-17</td>
<td>1.97E-17</td>
<td>-3.2</td>
<td>-3.14E-16</td>
<td>1.87E-16</td>
</tr>
<tr>
<td>13</td>
<td>2.63E-19</td>
<td>7.32E-20</td>
<td>3.6</td>
<td>-6.67E-19</td>
<td>1.19E-18</td>
</tr>
<tr>
<td>14</td>
<td>-4.77E-22</td>
<td>1.22E-22</td>
<td>-3.9</td>
<td>-2.02E-21</td>
<td>1.07E-21</td>
</tr>
</tbody>
</table>

Coefficients for the validated normative model of human ovarian volume throughout life. Each coefficient value is reported together with estimates of the standard error, T-statistic and 95% confidence limits for the value.

doi:10.1371/journal.pone.0071465.t001
and declines throughout life to about 2.8 mL (95% CI 2.7–2.9 mL) at the menopause.

The data do not support the notion of two distinct populations, PCOS and non-PCOS, giving a bimodal distribution of ovarian volumes at a given age. Model residual plots for ages up to 10 years are approximately normally distributed (Figure 5). Model residual plots for ages 10 through 30 years (Figure 6) and over 30 years (Figure 7) are close to an ideal normal distribution.

When the data is censored to remove 444 values over 10 mL, in line with the Rotterdam criteria for PCOS[7,8], the model changes slightly both in qualitative and quantitative terms, with a coefficient of determination \( r^2 \approx 0.69 \) for both models. The censored-data model is the same as the full-data model for young and old ages – average volume 0.7 mL (95% CI 0.4–1.1 mL) at 2 years and 2.8 mL (95% CI 2.7–2.9 mL) at age 50 years. The censored-data model has a lower peak predicted ovarian volume in the average case, 6.4 mL (95% CI 5.4–7.6 mL), with the peak occurring one year later at 21 years. This lower peak is outside the 95% confidence interval 6.5–9.2 mL for the full model peak, suggesting a statistically significant difference between the two peak values.

Discussion

We have described and validated the first normative model that describes ovarian volume in healthy females from conception to 82 years. The model has a coefficient of determination \( r^2 = 0.69 \) indicating that 69% of the variation in ovarian volumes throughout life is due to age alone. Ovarian volume rises through childhood and adolescence and is maximal in the average woman at 20 years of age, declining thereafter towards the menopause and beyond.

Transvaginal ultrasound evaluation has been used as an indirect assessment of ovarian reserve in adult sexually active females [16]. We have previously shown a strong positive correlation \( (r = 0.89) \) between NGF numbers and ovarian volume from ages 25 to 51 years [10], i.e. during the time that both are declining. Our normative model now adds to this by showing a steady rise in ovarian volume from birth (Figure 2) with a modest acceleration around the onset of puberty (age 9–10 years). The major contribution to ovarian volume before puberty is likely to be stromal growth; while small antral follicles are present in the ovaries of prepubertal girls of all ages [17], larger follicles are not found while serum gonadotrophin concentrations remain low. After menarche and the onset of ovulation the major contribution to changing ovarian volume is likely to be the number and size of the antral follicles present.

Human growth in childhood is described as three additive and partly superimposed components: infancy, childhood and puberty [18]. Each component appears to be controlled by distinct biological mechanisms. The infancy component is largely nutrition dependent, the childhood component is mostly dependent on growth hormone (GH) and the pubertal component depends on the synergism between sex steroids and GH. The slow rise in ovarian volume throughout mid-childhood (Figures 1 and 2) followed by an increase in ovarian volume during the pubertal years suggests that GH, in addition to sex steroids, may have an important role in determining ovarian size (and possibly function) in the early and late childhood years. A role for GH in determining ovarian size and volume during childhood and...
puberty is suggested by data from Bridges et al. 1993 who studied girls with growth disorders: GH insufficiency, skeletal dysplasia, and tall stature [13]. Total ovarian volume of untreated GH-insufficient girls was significantly less than that of GH-insufficient girls on GH treatment, girls with skeletal dysplasia on GH treatment, and girls with tall stature. They also found that tall girls had significantly greater ovarian volume than either of the GH-treated groups.

The measurement of ovarian volume has been found to be useful in a wide range of disorders in children and young females. Measurement of ovarian volume is an accurate diagnostic tool for adolescent girls with irregular menses. In the majority of these girls, enlarged ovaries are associated with polycystic ovary syndrome (PCOS) [19] and ovarian volume is part of the diagnostic criteria for that condition [7,8]. We therefore censored our dataset to exclude all women with ovarian volume greater than 10 mL. As the descriptions of subjects included in the original references varies, women with PCOS – or asymptomatic women whose ovaries had polycystic ovary morphology (PCOM) may have been included. In the largest data source, PCOS was not actively excluded: “...Patients with a solid or cystic ovarian tumor detected by sonography were excluded from this investigation since the purpose of this study was to determine normal ovarian volume...” [15]. Excluding these data points resulted in a reduction in the peak average ovarian volume, as would be expected, and a slight increase in the age at which the peak was reached. Importantly, our analysis does not address the validity of the criteria for the diagnosis of PCOS. Recent results suggest that antral follicle counts have better discriminatory performance than ovarian volume [20].

Girls with precocious puberty have significantly increased ovarian volumes compared with a normal population [21] and ovarian volume has been proposed as a useful discriminator between central precocious puberty and premature thelarche [22]. Furthermore, measurement of ovarian volume is a useful index with which to assess the efficacy of treatment of central precocious puberty with GnRH analogues [23].

The role of transvaginal USS as a screening test for ovarian cancer remains an important area of study [9,15,24] and transvaginal USS has an established role in the assessment and management of subfertility and in-vitro fertilization (IVF) in adult women [12,25]. It remains difficult to assess ovarian reserve in adolescents and young women with cancer due to the considerable age-related changes in the various markers available. The measurement of ovarian volume in addition to AMH may help predict which young women are at particular risk of premature ovarian insufficiency following cancer treatment and who may therefore benefit from fertility preservation techniques [26,27].

Our model is derived from data from multiple sources of the measurement of ovarian volume in otherwise healthy females. This is both a strength and a weakness of the study. The strength is that the measuring errors, both underestimating and overestimating ovarian volume, are likely to be negated as any bias is unlikely to be always in the same direction for each data source. The weakness is the heterogeneity of the values obtained from diverse sources. We cannot be certain that the measurement of ovarian volume by abdominal ultrasound, which is often difficult in young
children, is as accurate as measurement by transvaginal ultrasound in older females [28]. Similarly measurements taken at MRI may be different from those obtained by weighing the ovary following oophorectomy and calculating the volume from weight. The largest data source consists of values imputed from a very large data source obtained by transvaginal ultrasound as part of a screening programme for ovarian cancer [15]. This study excluded patients with a solid or cystic ovarian tumor detected by sonography, but not patients with polycystic ovary morphology. Our normative model of ovarian volume using data derived from multiple data sources and different methods of assessment overcomes the weakness of other studies in which only one imaging modality is used, because any potential bias in one direction is likely to be negated.

We have shown that in the average case ovarian volume rises from 0.7 mL (95% CI 0.4–1.1 mL) at 2 years of age to a peak of 7.7 (95% CI 6.5–9.2 mL) mL at 20 years of age and declines throughout later life to about 2.8 mL (95% CI 2.7–2.9 mL) at the menopause. This is the first validated normative model of ovarian volume from conception to old age; it will be of use in the diagnosis and management of a number of diverse gynaecological and reproductive conditions in females from birth to menopause and beyond.

**Materials and Methods**

The research methodology used both for data acquisition and data analysis closely follows that used to derive a validated normative model for the level of anti-Müllerian hormone (AMH) found in the blood of healthy human females for ages from conception to menopause [29,30].

**Ethics statement**

Permission to perform basic science studies on the ovarian material retrieved in Denmark was given by the Minister of Health in Denmark and by the Committee on Biomedical Research Ethics of the Capital Region on 21st September 2011 (protocol number H-2-2011-044). Written informed consent for the original human work that produced the tissue samples was obtained, and all data were anonymised prior to analysis.

**Data acquisition**

The data for this study (Table 3) come from three sources: our own measurements of ovarian volume, imputation from the large-scale study by Pavlik et al. [15] as described in [10], and publications in the scientific literature. Taken as a single dataset, it approximates the healthy population in terms of ovarian volume, for ages ranging from mid-term fetal to postmenopausal.

We included data from two unpublished sources. Firstly, a detailed assessment of 300 MRI examinations in children without known endocrine, chromosomal or oncological conditions that included the pelvis, yielded 49 pairs of ovaries where both ovaries were visualized and measurable in three dimensions (median age 13 years, range 2 to 16.7 years). Ovarian volumes were calculated using the prolate ellipsoid approximation formula $a \times b \times c \times \frac{a}{2}$. Secondly, a further 384 ovaries (median age 27.5 years, range 0.5 to 39.8 years) were weighed before cryopreservation at the University Hospital of Copenhagen. Subjects were known to have...
non-ovarian cancer; subjects who had received chemotherapy were excluded. Ovarian volume was estimated using the published conversion factor for ovarian tissue density: 1.00 g/mL [31].

Summary statistics were extracted from Pavlik et al. [15] for ages 24–85 years. Repeated (10-fold) parametric bootstrapping [32] was used to simulate datapoints from the published distributions to obtain a single dataset (\(n \sim 58,255\)) that accurately reproduces the published results.

In order to obtain data from the existing literature with emphasis on volumes earlier in life than the 24 years minimum age reported in Pavlik et al. [15] studies of ovarian volume in normal, healthy girls were identified using Medline and PubMed searches using the search terms Ovary, Child, Ovarian size/volume, Normal, Healthy and Neonatal. The references of these identified studies were then reviewed, and any other relevant research papers were extracted. Papers were included if they contained ovarian volume results for healthy, normal girls with no ovarian or endocrinological abnormalities, so as to isolate data that approximate the healthy human population. Abstracts of 37 studies were identified via this method.

After analysis of the full papers, studies were excluded if either (i) the results consisted purely of descriptive statistics, or (ii) subjects were classified by pubertal stage rather than age. Of the remaining nine studies, seven contained data measured by trans-abdominal ultrasound and plotted in graphs [14,33–38] while two contained tabular data (with fetal/neonatal ovaries extracted and measured/sliced to calculate volumes) [39,40]. The data was extracted from the graphs (\(n = 1,151\)) using Plot Digitizer software [41], and combined with the tabular data (\(n = 64\)). Ovarian volumes were standardised to the prolate ellipsoid approximation formula \(a \times b \times c \times \frac{2}{\pi}\) since some studies used the variation \(a \times b \times c \times \frac{1}{2}\).

### Data analysis

Zero volume values at conception were added to the combined dataset (Table 3), in order to force models through the only known volume at any age. Since variability increases with ovarian volume, we log-adjusted the data (after adding one to each value so that zero volume on a chart represents zero ovarian volume). We then fitted 310 mathematical models to the training data using TableCurve-2D (Systat Software Inc., San Jose, California, USA), and ranked the results by coefficient of determination, \(r^2\). Each model defines a generic type of curve and has parameters which, when instantiated gives a specific curve of that type. For each model we calculated values for the parameters that maximise the \(r^2\) coefficient. The Levenberg-Marquardt non-linear curve-fitting algorithm was used throughout, with convergence to 6 significant figures after a maximum of 1,500 iterations. For each candidate
Figure 5. Model residuals for ages up to 10 years. Residuals are the squared differences between data values and predicted values for that age.
doi:10.1371/journal.pone.0071465.g005

Figure 6. Model residuals for ages between 10 and 30 years. Residuals are the squared differences between data values and predicted values for that age.
doi:10.1371/journal.pone.0071465.g006
model, the mean square error and $r^2$ were calculated after removing the artificial zero values at conception.

The best performing family of models were high precision polynomials. 5-fold cross validation was performed: the data were randomly split into 5 equally sized subsets. For each subset $S$, the other four subsets were used to train high precision polynomials of degree 8 through 20, with subset $S$ being held back as test data. The mean square error of the test data was calculated and compared to the mean square error of training data for the same model. In other words, the estimated prediction error of a model when generalized to unseen data was compared to the training error of the model. A model was considered validated if

1. the residuals of the test data were approximately normally distributed (Figure 3); and
2. the tradeoff between high $r^2$ (denoting possible overfitting to the data) and low generalisation error (denoting possible underfitting to the data) was optimal (Figure 4).

We tested for bimodal volume distributions that would suggest distinct PCOS and non-PCOS sub-populations by analysis of model residuals for age ranges up to 10 years, 10–30 years, and above 30 years. Normally-distributed residuals for log-adjusted values correspond with skew-normal population volumes (i.e. a single population with PCOS and non-PCOS volumes forming a smooth continuum of values). Significant variances from normality provide evidence for a distinct PCOS sub-population.

The validated model was also assessed against the Rotterdam criteria for PCOS [7,8] by censoring all values above the 10 mL discriminatory cutoff volume, re-fitting the model, and comparing peak ages and volumes.

Table 3. Ovarian volume data sources.

<table>
<thead>
<tr>
<th>Study</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref.</td>
<td>First author</td>
</tr>
<tr>
<td>[10]</td>
<td>Kelsey</td>
</tr>
<tr>
<td>[33]</td>
<td>Badouraki</td>
</tr>
<tr>
<td>[34]</td>
<td>azzaghy-Azar</td>
</tr>
<tr>
<td>[35]</td>
<td>Seth</td>
</tr>
<tr>
<td>[36]</td>
<td>Holm</td>
</tr>
<tr>
<td>[37]</td>
<td>Ziereisen</td>
</tr>
<tr>
<td>[38]</td>
<td>Griffin</td>
</tr>
<tr>
<td>[14]</td>
<td>Stanhope</td>
</tr>
<tr>
<td>*</td>
<td>Wilkinson</td>
</tr>
<tr>
<td>*</td>
<td>Andersen</td>
</tr>
<tr>
<td>[39]</td>
<td>Sforza</td>
</tr>
<tr>
<td>[40]</td>
<td>Simkins</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
</tr>
</tbody>
</table>

The year column refers to the year of publication; * denotes our own unpublished data.

doi:10.1371/journal.pone.0071465.t003

Figure 7. Model residuals for ages over 30 years. Residuals are the squared differences between data values and predicted values for that age. doi:10.1371/journal.pone.0071465.g007
Author Contributions
Conceived and designed the experiments: TWK SKD AGW TG CYA RAA WHBW. Performed the experiments: TWK SKD AGW TG CYA.

References